

SYSTEMS AND METHODS FOR TRANSLATING CHINESE PINYIN TO CHINESE CHARACTERS

ABSTRACT OF THE DISCLOSURE

Systems and methods to process and translate pinyin to Chinese characters and
5 words are disclosed. A Chinese language model is trained by extracting unknown
character strings from Chinese inputs, e.g., documents and/or user inputs/queries,
determining valid words from the unknown character strings, and generating a transition
matrix based on the Chinese inputs for predicting a word string given the context. A
method for translating a pinyin input generally includes generating a set of Chinese
10 character strings from the pinyin input using a Chinese dictionary including words
derived from the Chinese inputs and a language model trained based on the Chinese
inputs, each character string having a weight indicating the likelihood that the character
string corresponds to the pinyin input. An ambiguous user input may be classified as
non-pinyin or pinyin by identifying an ambiguous pinyin/non-pinyin ASCII word in the
15 user input and analyzing the context to classify the user input.